ON THE EXPECTED BETTI NUMBERS OF THE NODAL SET OF RANDOM FIELDS

IGOR WIGMAN

Abstract. This note concerns the asymptotics of the expected total Betti numbers of the nodal set for an important class of Gaussian ensembles of random fields on Riemannian manifolds. By working with the limit random field defined on the Euclidean space we were able to obtain a locally precise asymptotic result, though due to the possible positive contribution of large *percolating* components this does not allow to infer a global result. As a by-product of our analysis, we refine the lower bound of Gayet-Welschinger for the important Kostlan ensemble of random polynomials and its generalisation to Kähler manifolds.

1. Introduction

1.1. **Betti numbers for random fields: Euclidean case.** Let $F:\mathbb{R}^d\to\mathbb{R}$ be a centred stationary Gaussian random field, $d\ge 2$. The *nodal set of* F is its (random) zero set

$$\mathcal{Z}_F := F^{-1}(0)=\{x\in\mathbb{R}^d:\ F(x)=0\}\subseteq\mathbb{R}^d;$$

assuming F is sufficiently *smooth* and *non-degenerate* (or regular), its connected components ("nodal components of F") are a.s. either closed $(d-1)$-manifolds or smooth infinite hypersurfaces ("percolating components"). One way to study the topology of $\mathcal{Z}_F$, a central research thread in the recent few years, is by restricting F to a large centred ball $B(R)=\{x\in\mathbb{R}^d:\ \|x\|<R\}$, and then investigate the restricted nodal set $\widetilde{\mathcal{Z}_F}(R):=F^{-1}(0)\cap B(R)$ as $R\to\infty$. The set $\widetilde{\mathcal{Z}_F}(R)$ consists of the union of the a.s. smooth closed nodal components of $\mathcal{Z}_F$ lying entirely in $B(R)$, and the fractions of nodal components of F intersecting $\partial B(R)$; note that, by intersecting with $B(R)$, the components intersecting $\partial B(R)$, finite or percolating, might break into 2 or more connected components, or fail to be closed.

It follows as a by-product of the precise analysis due to Nazarov-Sodin [27, 20] that, under very mild assumptions on F to be discussed below, mainly concerning its smoothness and non-degeneracy, with high probability *most* of the components of $\mathcal{Z}_F$ fall into the former, rather than the latter, category (see (1.2) below). That is, for R large, with high probability, most of the components of $\mathcal{Z}_F$ intersecting $B(R)$ are lying entirely within $B(R)$. Setting

$$\mathcal{Z}_F(R):=\bigcup_{\gamma\subseteq B(R)}\gamma$$

to be the union of all the nodal components γ of F lying entirely in $B(R)$, the first primary concern of this note is in the topology of $\mathcal{Z}_F(R)$, and, in particular, the Betti numbers of $\mathcal{Z}_F(R)$ as $R\to\infty$, more precisely, the asymptotics of their expected values.

Date: November 18, 2019.

For $0 \le i \le d-1$ the corresponding Betti number $b_i(\cdot)$ is the dimension of the i'th homology group, so that a.s.

$$\beta_i(R) = \beta_{F;i}(R) := b_i(\mathcal{Z}_F(R)) = \sum_{\gamma \subseteq \mathcal{Z}_F(R)} b_i(\gamma), \tag{1.1}$$

summation over all nodal components γ lying in $\mathcal{Z}_F(R)$. For example, $\beta_0 =: \mathcal{N}_F(R)$ is the total number of connected components $\gamma \subseteq \mathcal{Z}_F(R)$ ("nodal count") analysed by Nazarov-Sodin, and

$$\beta_i(R) = \beta_{d-1-i}(R)$$

by Poincaré duality. To be able to state Nazarov-Sodin's results we need to introduce the following axioms; by convention they are expressed in terms of the spectral measure rather than F or its covariance function.

Definition 1.1 (Axioms $(\rho 1) - (\rho 4)$ on F)**.** Let $F : \mathbb{R}^d \to \mathbb{R}$ be a Gaussian stationary random field,

$$r_F(x-y) = r_F(x,y) := \mathbb{E}[F(x) \cdot F(y)]$$

the covariance function of F, and $\rho = \rho_F$ be its spectral measure, i.e. the Fourier transform of r_F on $\mathbb{R}^d$.

(1) F satisfies $(\rho 1)$ if the measure ρ has no atoms.
(2) F satisfies $(\rho 2)$ if for some $p > 6$,

$$\int_{\mathbb{R}^d} \|\lambda\|^p d\rho(\lambda) < \infty.$$

(3) F satisfies $(\rho 3)$ if the support of ρ does not lie in a linear hyperplane of $\mathbb{R}^d$.
(4) F satisfies $(\rho 4)$ if the interior of the support of ρ is non-empty.

Axioms $(\rho 1)$, $(\rho 2)$ and $(\rho 3)$ ensure that the action of translations on $\mathbb{R}^d$ is ergodic, a.s. sufficient smoothness of F, and non-degeneracy of F understood in proper sense, respectively. Axiom $(\rho 4)$ implies that any smooth function belongs to the support of the law of F, which, in turn, will yield the positivity of the number of nodal components, and positive representation of every topological type of nodal components.

Recall that $\mathcal{N}_F(R) = \beta_0(R)$ is the number of nodal components of F entirely lying in $B(R)$, and let V_d be the volume of the unit d-ball, and $\operatorname{Vol} B(R) = V_d \cdot R^d$ be the volume of the radius R ball in $\mathbb{R}^d$. Nazarov and Sodin [27, 20] proved that if F satisfies $(\rho 1) - (\rho 3)$, then there exists a constant $c_{NS} = c_{NS}(\rho_F)$ ("Nazarov-Sodin constant") so that $\frac{\mathcal{N}_F(R)}{\operatorname{Vol} B(R)}$ converges to c_{NS}, both in mean and a.s. That is, as $R \to \infty$,

$$\mathbb{E}\left[\left|\frac{\mathcal{N}_F(R)}{\operatorname{Vol} B(R)} - c_{NS}\right|\right] \to 0, \tag{1.2}$$

so that, in particular,

$$\mathbb{E}[\mathcal{N}_F(R)] = c_{NS} \cdot \operatorname{Vol} B(R) + o(R^d). \tag{1.3}$$

They also showed that imposing $(\rho 4)$ is sufficient (but not necessary) for the strict positivity of c_{NS}, and found other very mild sufficient conditions on ρ, so that $c_{NS} > 0$. The validity of the asymptotic (1.3) for the expected nodal count was extended [16] to hold without imposing the ergodicity axiom ($\rho 1$), with $c_{NS} = c_{NS}(\rho_F)$ appropriately generalised, also establishing a stronger estimate for the error term as compared to the r.h.s. of (1.3).

One might think that endowing the "larger" components with the same weight 1 as the "smaller" components might be "discriminatory" towards the larger ones, so that separating the counts based on the components' topology [26, 25] or geometry [4] would provide an adequate response for the alleged discrimination. These nevertheless do not address the important question of the *total* Betti number β_i, the main difficulty being that the individual Betti number $b_i(\gamma)$ of a nodal component γ of F is not bounded, even under the assumption that $\gamma \subseteq B(R)$ is entirely lying inside a compact domain. Despite this, we will be able to resolve this difficulty by controlling from above the total Betti number via Morse Theory [18], an approach already pursued by Gayet-Welschinger [11] (see §2 below for a more detailed explanation).

Theorem 1.2. *Let $F : \mathbb{R}^d \to \mathbb{R}$ be a centred Gaussian random field, satisfying axioms ($\rho 2$) and ($\rho 3$) of Definition 1.1, $d \geq 2$, and $0 \leq i \leq d-1$. Then*

a. There exists a number $c_i = c_{F;i} \geq 0$ so that

$$\mathbb{E}[\beta_i(R)] = c_i \cdot \operatorname{Vol} B(R) + o_{R\to\infty}(R^d). \tag{1.4}$$

b. If, in addition, F satisfies ($\rho 1$), then convergence (1.4) could be extended to hold in mean, i.e.

$$\mathbb{E}\left[\left|\frac{\beta_i(R)}{\operatorname{Vol} B(R)} - c_i\right|\right] \to 0 \tag{1.5}$$

as $R \to \infty$.

c. Further, if F satisfies the axiom ($\rho 4$) (in addition to ($\rho 2$) and ($\rho 3$), but not ($\rho 1$)), then $c_i > 0$. The same conclusion holds for the important Berry's monochromatic isotropic random waves in arbitrary dimensions ("Berry's random wave model").

1.2. **Motivation and background.** The Betti numbers of both the nodal and the excursion sets of Gaussian random fields serve as their important topological descriptor, and are therefore addressed in both mathematics and experimental physics literature, in particular cosmology [23]. From the complex geometry perspective Gayet and Welschinger [11] studied the distribution of the total Betti numbers of the zero set for the Kostlan Gaussian ensemble of degree n random homogeneous polynomials on the d-dimensional projective space, and their generalisation to Kähler manifolds, $n \to \infty$. In the projective coordinates $x = [x_0 : \ldots : x_d] \in \mathbb{R}\mathcal{P}^d$ we may write

$$P_n(x) = \sum_{|j|=n} \sqrt{\binom{n}{j}} a_j x^j, \tag{1.6}$$

where $j = (j_0, \ldots, j_d)$, $|j| = \sum\limits_{i=0}^{d} j_i$, $x^j = x_0^{j_0} \cdot \ldots \cdot x_d^{j_d}$, $\binom{n}{j} = \frac{n!}{j_0! \cdot \ldots \cdot j_d!}$, and $\{a_j\}$ are standard Gaussian i.i.d. By the homogeneity of P_n, its zero set makes sense on the projective space. The Kostlan (also referred to as "Shub-Smale") ensemble is an important model of random polynomials, uniquely invariant w.r.t. unitary transformations on $\mathbb{CP}^d$. Restricted to the unit sphere $\mathcal{S}^d \subseteq \mathbb{R}^{d+1}$, the random fields P_n are defined by the covariance function

$$\mathbb{E}[P_n(x) \cdot P_n(y)] = \langle x, y \rangle^n = (\cos(\theta(x, y)))^n ,$$

where $x, y \in \mathcal{S}^d$, the inner product $\langle \cdot, \cdot \rangle$ is inherited from $\mathbb{R}^{d+1}$, and $\theta(\cdot, \cdot)$ is the angle between two points on $\mathbb{R}^{d+1}$.

Upon scaling by $\sqrt{n}$ (the meaning is explained in Definition 1.3 below), the Kostlan polynomials (1.6) admit [27, §2.5.4], locally uniformly, a (stationary isotropic) limit random field on $\mathbb{R}^d$, namely the Bargmann-Fock ensemble defined by the "Gaussian" covariance kernel

$$\kappa(x) := e^{-x^2/2}, \tag{1.7}$$

see also [3, 5]. This indicates that one should expect the Betti numbers to be of order of magnitude $\approx n^{d/2}$. That this is so is supported by Gayet-Welschinger's upper bounds [11]

$$\mathbb{E}[b_i(P_n^{-1}(0))] \le A_i n^{d/2}$$

with some semi-explicit $A_i > 0$, and the subsequent lower bounds [10]

$$\mathbb{E}[b_i(P_n^{-1}(0))] \ge a_i n^{d/2}, \tag{1.8}$$

$a_i > 0$, but to our best knowledge the important question of the true asymptotic behaviour of $\mathbb{E}[b_i(P_n^{-1}(0))]$ is still open.

1.3. **Betti numbers for Gaussian ensembles on Riemannian manifolds.** Since κ of (1.7) (or, rather, its Fourier transform) easily satisfies all Nazarov-Sodin's axioms $(\rho 1) - (\rho 4)$ of Definition 1.1, one wishes to invoke Theorem 1.2 with the Bargmann-Fock field in place of F, and try to deduce the results analogous to (1.4) for the Betti numbers of the nodal set of P_n in (1.6). This is precisely the purpose of Theorem 1.5 below, valid in a scenario of *local translation invariant limits*, far more general than merely the Kostlan ensemble, whose introduction is our next goal.

Let $\mathcal{M}$ be a compact Riemannian d-manifold, and $\{f_L\}_{L\in\mathcal{L}}$ be a family of *smooth* Gaussian random fields $f_L : \mathcal{M} \to \mathbb{R}$, where the index L attains a *discrete* set $\mathcal{L}$, and $K_L(\cdot, \cdot)$ the covariance function corresponding to f_L, so that

$$K_L(x, y) = \mathbb{E}[f_L(x) \cdot f_L(y)];$$

the parameter L should be thought of as the scaling factor, generalising the role of $\sqrt{n}$ for the Kostlan ensemble. We scale f_L restricted to a sufficiently small neighbourhood of a point $x \in \mathcal{M}$, so that the exponential map $\exp_x(\cdot) : T_x\mathcal{M} \to \mathcal{M}$ is well defined. We define

$$f_{x,L}(u) := f_L(\exp_x(u/L), \tag{1.9}$$

with covariance

$$K_{x,L}(u, v) := K_L(\exp_x(u/L), \exp_x(v/L))$$

with $|u|, |v| < L \cdot r$ with r sufficiently small, uniformly with $x \in \mathcal{M}$, allowing u, v to grow with $L \to \infty$.

Definition 1.3 (Local translation invariant limits, cf. [20, Definition 2 on p. 6])**.** We say that the Gaussian ensemble $\{f_L\}_{L\in\mathcal{L}}$ possesses local translation invariant limits, if for almost all $x\in\mathcal{M}$ there exists a positive definite function $K_x:\mathbb{R}^d\to\mathbb{R}$, so that for all $R>0$,

$$\lim_{L\to\infty}\sup_{|u|,|v|\le R}|K_{x,L}(u,v)-K_x(u-v)|\to 0. \tag{1.10}$$

Important examples of Gaussian ensembles possessing translation invariant local limits include (but not limited to) Kostlan's ensemble (1.6) of random homogeneous polynomials, and Gaussian band-limited functions [25], i.e. Gaussian superpositions of Laplace eigenfunctions corresponding to eigenvalues lying in an energy window. For manifolds with spectral degeneracy, such as the sphere and the torus (and d-cube with boundary), the *monochromatic* random waves (i.e. Gaussian superpositions of Laplace eigenfunctions belonging to the same eigenspace) are a particular case of band-limited functions; two of the most interesting cases are those of random spherical harmonics (random Laplace eigenfunctions on the round unit d-sphere) [28, 29], and "Arithmetic Random Waves" (random Laplace eigenfunctions on the standard d-torus) [22, 15].

In all the said examples of Gaussian ensembles on manifolds of our particular interest the scaling limit K_x (and the associate Gaussian random field on $\mathbb{R}^d$) was independent of x, and the limit in (1.10) is uniform, attained in a strong quantitative form, see the discussion in [6, §2.1]. We will also need the following, more technical concepts of uniform smoothness and non-degeneracy for $\{f_L\}$, introduced in [27, definitions 2-3, p. 14-15].

Definition 1.4 (Smoothness and non-degeneracy)**.**

(1) We say that $\{f_L\}$ is C^{3-} smooth if for every $0<R<\infty$,
$$\limsup_{L\to\infty}\sup\left\{|\partial_u^i\partial_v^j K_{x,L}(u,v)|:\ |i|,|j|\le 3;\ x\in\mathcal{M},\|u\|,\|v\|\le R\right\}<\infty. \tag{1.11}$$

(2) We say that $\{f_L\}$ is non-degenerate if for every $0<R<\infty$
$$\liminf_{L\to\infty}\inf\left\{\mathbb{E}\left[\partial_\xi f_{x,L}(u)^2\right]:\ \xi\in\mathcal{S}^{d-1},\,x\in\mathcal{M},\,\|u\|\le R\right\}>0.$$

Let $\{f_L\}_{L\in\mathcal{L}}$ be a C^{3-} smooth, non-degenerate, Gaussian ensemble possessing translation invariant local limits K_x, corresponding to Gaussian random fields on R^d with spectral measure ρ_x, satisfying axioms $(\rho 1)-(\rho 3)$. Denote $\mathcal{N}(f_L;x,R/L)$ to be the number of nodal components of f_L lying entirely in the geodesic ball $B_x(R/L)\subseteq\mathcal{M}$, and $\mathcal{N}(f_L)$ to be the *total* number of the nodal components of f_L on $\mathcal{M}$. In this settings Nazarov-Sodin [27, 20] proved that

$$\lim_{R\to\infty}\limsup_{L\to\infty}\mathbb{E}\left[\left|\frac{\mathcal{N}(f_L;x,R/L)}{\operatorname{Vol}B(R)}-c_{NS}(\rho_x)\right|\right]=0, \tag{1.12}$$

with $c_{NS}(\cdot)$ same as in (1.2).

For the total number $\mathcal{N}(f_L)$ they glued the local results (1.12), to deduce, on invoking a two-parameter analogue of Egorov's Theorem yielding the *almost uniform* convergence of (1.12) w.r.t. x, that

$$\lim_{L\to\infty}\mathbb{E}\left[\left|\frac{\mathcal{N}(f_L)}{V_dL^d}-\nu\right|\right]\to 0, \tag{1.13}$$

holds with

$$\nu := \int_{\mathcal{M}} c_{NS}(\rho_x)dx.$$

In particular, (1.13) yields

$$\mathbb{E}[\mathcal{N}(f_L)] = V_d\nu \cdot L^d + o(L^d), \tag{1.14}$$

As it was mentioned above, in practice, in many applications, the scaling limit $K_x(\cdot) \equiv K(\cdot)$ does not depend on x, so that, assuming w.l.o.g. that $\mathrm{Vol}(\mathcal{M}) = 1$, the asymptotic constant ν in (1.13) (and (1.14)) is $\nu = c_{NS}(\rho)$, where ρ is the Fourier transform of K. In this situation, in accordance with Theorem 1.2c, $\nu = c_{NS} > 0$ is positive, if (ρ4) is satisfied. The following result extends (1.12) to arbitrary Betti numbers.

Theorem 1.5. *Let $\{f_L\}_{L\in\mathcal{L}}$ be a C^{3-} smooth, non-degenerate, Gaussian ensemble, $x \in \mathcal{M}$ satisfying* (1.10) *with some K_x satisfying axioms $(\rho 1) - (\rho 3)$, and $0 \le i \le d-1$. Denote $\beta_{i;L}(x, R/L) = \beta_i(f_L; x, R/L)$ to be the total i'th Betti number of the union of all components of $f_L^{-1}(0)$ entirely contained in the geodesic ball $B_x(R/L)$. Then for every $\epsilon > 0$*

$$\lim_{R\to\infty} \limsup_{L\to\infty} \mathcal{P}r\left\{ \left| \frac{\beta_{i;L}(x, R/L)}{\mathrm{Vol}\, B(R)} - c_i \right| > \epsilon \right\} = 0. \tag{1.15}$$

where c_i is the same as in (1.5)*, corresponding to the random field defined by K_x.*

Theorem 1.5 asserts that the random variables $\left\{\frac{\beta_{i;L}(x,R/L)}{\mathrm{Vol}\, B(R)}\right\}_{L\in\mathcal{L}}$ converge in probability to c_i, in the double limit $L \to \infty$, and then $R \to \infty$. One would be tempted to try to deduce the convergence in mean for the same setting, the main obstacle being that $\beta_{i;L}(x, R/L)$ is not bounded, and, in principle, a small probability event might contribute positively to the expectation of $\beta_{i;L}(x, R/L)$. While it is plausible (if not likely) that a handy bound on the variance (or the second moment), such as [8, 19], for the critical points number would rule this out and establish the desired L^1-convergence in this, or, perhaps, slightly more restrictive scenario, we will not pursue this direction in the present manuscript, for the sake of keeping it compact.

Theorem 1.5 applied on the Kostlan ensemble (1.6) of random polynomials, in particular, recovers Gayet-Welschinger's later lower bound (1.8), but, finer, with high probability, it prescribes the asymptotics of the total Betti numbers of all the components lying in geodesic balls of radius slightly above $1/\sqrt{n}$, and hence, in this case, one might think of Theorem 1.5 as a refinement of (1.8). It would be desirable to determine the true asymptotic law of $\mathbb{E}[b_i(P_n^{-1}(0))]$ (hopefully, for the more general scenario), though the possibility of giant ("percolating") components is a genuine consideration, and, if our present understanding of this subtlety is correct [6], then, to resolve the asymptotics of $\mathbb{E}[b_i(P_n^{-1}(0))]$ the question whether they consume a positive proportion of the total Betti numbers cannot be possibly avoided. In fact, it is likely that for $d \ge 3$, with high probability, there exists a single percolating component consuming a high proportion of the space, and contributing positively to the Betti numbers, as found numerically by Barnett-Jin (presented within [25]), and explained by P. Sarnak [24], with the use of percolating vs. non-percolating random fields (see [6, §1.2] for more details, and also the discussion in §2 below).

1.4. **Acknowledgement.** It is a pleasure to thank P. Sarnak and M. Sodin for their comments on the presented proofs of the main results, D. Panov for freely sharing his expertise on various aspects of the presented material, Z. Rudnick for his support and encouragement regarding this work and his valuable comments on an earlier version of this manuscript, and Z. Kabluchko for pointing out the superadditive ergodic theorem [14, Theorem 2.14, page 210]. The author of this manuscript is grateful to D. Beliaev and S. Muirhead for many stimulating conversations concerning subjects of high relelvance to the presented research. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013), ERC grant agreement n^{o} 335141.

2. Outline of the proofs and discussion

2.1. **Outline of the proofs of the principal results.** The principal novel result of this manuscript is Theorem 1.2. Theorem 1.2 *given*, the proof of Theorem 1.5 does not differ significantly from the proof of [27, Theorem 5] given [27, Theorem 1]. The key observation here is that while passing from the Euclidean random field F_x to its perturbed Riemannian version $f_{x,L}$ in the vicinity of $x \in \mathcal{M}$, the topology of its nodal set is preserved on a high probability *stable* event, to be constructed, and hence so is its i'th Betti number. In fact, this was the conclusion from the argument presented in [25, Theorem 6.2] that will be reconstructed in §4, alas briefly, for the sake of completeness.

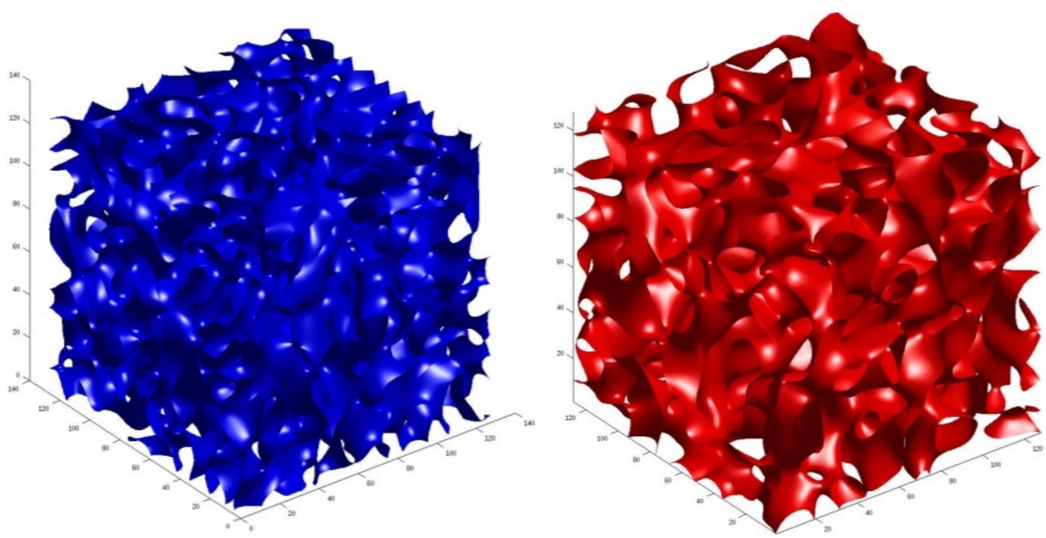

Figure 1. Computer simulations by A. Barnett. Left: Giant percolating nodal components for 3-dimensional monochromatic isotropic waves. Right: Analogous picture for the "Real Fubini-Study" (a random ensemble of homogeneous polynomials, with different law as compared to Kostlan's ensemble).

To address the asymptotic expected nodal count $\mathcal{N}_F(R) = \beta_{F;0}(R)$, Nazarov-Sodin have developed the so-called *Integral Geometric sandwich*. The idea is that one bounds $\mathcal{N}_F(R)$ from below using $\mathcal{N}_{\cdot}(r)$, of radii $0 < r < R$ much smaller than R ("fixed"), and F translated (equivalently, shifter radius-r ball), and from above using a version of $\mathcal{N}_F(r)$, where, rather than counting nodal components lying entirely in $B(r)$ (or its shift), we also include those components intersecting its boundary $\partial B(r)$. By invoking ergodic methods one shows that both these bounds converge to the same limit, and in its turn this yields automatically both the asymptotics for the expected nodal count, and the convergence in mean.

Unfortunately, since we endow each nodal component γ with the, possibly unbounded, weight $b_i(\gamma)$, the upper bound in the sandwich does not seemingly yield a useful result. We bypass this major obstacle by using a global bound on the expected Betti numbers via Morse Theory (and the Kac-Rice method), and then establishing an asymptotics for the expected number. Rather than working with arbitrary chosen "fixed" radii $r > 0$, we only work with "good" radii, defined so that these numbers are "almost maximising" the expected Betti numbers, so that we can infer the same for all the sufficiently big radii $R > r$ (see (3.11) and (3.13)). In hindsight, we interpret working with the good radii as "miraculously" eliminating the possible fluctuations in the contribution to the Betti numbers of the giant percolating domains. Once the asymptotics for the expected Betti number has been determined, we tour de force working with the good radii to also yield the convergence in mean, with the help of the ergodic assumption ($\rho 1$).

Another possible strategy for proving results like Theorem 1.2 is by observing that, by naturally extending the definition of β_i to smooth domains $\mathcal{D} \subseteq \mathbb{R}^d$ as

$$\beta_i(\mathcal{D}) = \beta_{i;F}(\mathcal{D}) := \sum_{\gamma \subseteq \mathcal{D}} b_i(\gamma),$$

with summation over the (random) nodal components of F lying in $\mathcal{D}$, $\beta_i(\cdot)$ is made into a *super-additive random variable*, i.e. for all $\mathcal{D}_1, \ldots, \mathcal{D}_k \subseteq \mathbb{R}^d$ pairwise disjoint domains, the inequality

$$\beta_i\left(\bigcup_{j=1}^{k} \mathcal{D}_j\right) \geq \sum_{j=1}^{k} \beta_i(\mathcal{D}_j)$$

holds. It then might be tempting to apply the superadditive ergodic theorem [14, Theorem 2.14, page 210] (and its finer version [21, p. 165]) on β_i. However, in this manuscript we will present a direct and explicit treatise of this subject.

2.2. **Discussion.** As it was mentioned above, a straightforward application of 1.5 on the Kostlan's ensemble of random homogenous polynomials, in particular implies the lower bound (1.8) for the total expected Betti number for this ensemble due to Gayet-Welschinger, and its generalisations for Kähler manifolds. Our argument is entirely different as compared to Gayet-Welschinger's: rather than working with the finite degree polynomials (1.6), as in [10], we first prove the result for the limit Bargmann-Fock random field on $\mathbb{R}^d$ (Theorem 1.2), and then deduce the result by a perturbative procedure following Nazarov-Sodin (Theorem 1.5).

It is crucial to determine whether the global asymptotics

$$\mathbb{E}[\beta_{i;L}] \sim c_i \operatorname{Vol}(\mathcal{M}) \cdot L^d,$$

expected from its local probabilistic version (1.15), could be extended to hold for the total expected Betti number of $f^{-1}(0)$ in some scenario, inclusive of all the motivational examples. Such a result would indicate that no giant "percolating" components, not lying inside any *macroscopic* (or slightly bigger) geodesic balls exist, contributing positively to the Betti numbers. In fact some numerics due to Barnett-Jin (presented within [25]) support the contrary for $d \geq 3$, as argued by Sarnak [24], see Figure 1, and also [6, §2.1]. To our best knowledge, at this stage this question is entirely open, save for the results on $\beta_{0;L}$ (and $\beta_{d-1;L}$) due to Nazarov-Sodin.

3. Proof of Theorem 1.2

3.1. **Auxiliary lemmas.** Recall that $\beta_i(R) = b_i(\mathcal{Z}_F(R))$ is defined in (1.1), and for $x \in \mathbb{R}^d$, $R > 0$, introduce

$$\beta_i(x; R) = \beta_{F;i}(x, R) := \sum_{\gamma \subseteq \mathcal{Z}_F \cap B_x(R)} b_i(\gamma), \tag{3.1}$$

summation over all nodal components of F contained in the shifted ball $B_x(R)$, or, equivalently

$$\beta_{F;i}(x, R) = \beta_{T_x F;i}(R),$$

where T_x acts by translation $(T_x F)(\cdot) = F(\cdot - x)$.

Lemma 3.1 (Integral-Geometric sandwich, lower bound; cf. [27, Lemma 1])**.** *For every $0 < r < R$ we have the following inequality*

$$\frac{1}{\operatorname{Vol} B(r)} \int\limits_{B(R-r)} \beta_i(x; r) dx \leq \beta_i(R). \tag{3.2}$$

Proof. Since if a nodal component of F is contained in $B_x(r)$ for some $x \in B(R-r)$, then $\gamma \subseteq B(R)$, we may invert the order of summation and integration to write:

$$\begin{aligned}
\frac{1}{\operatorname{Vol} B(r)} \int\limits_{B(R-r)} \beta_i(x; r) dx &= \frac{1}{\operatorname{Vol} B(r)} \int\limits_{B(R-r)} \sum_{\gamma \subseteq \mathcal{Z}(F)} \mathbb{1}_{\gamma \subseteq B_x(r)} \cdot b_i(\gamma) dx \\
&= \frac{1}{\operatorname{Vol} B(r)} \sum_{\gamma \subseteq \mathcal{Z}(F) \cap B(R)} b_i(\gamma) \cdot \operatorname{Vol}\{x \in B(R-r) : \ \gamma \subseteq B_x(r)\} \\
&\leq \sum_{\gamma \subseteq \mathcal{Z}(F) \cap B(R)} b_i(\gamma) = b_i(R),
\end{aligned}$$

since

$$\{x \in B(R-r) : \ \gamma \subseteq B_x(r)\} = \bigcap_{y \in \gamma} B_y(r)$$

is of volume $\leq \operatorname{Vol} B(r)$. □

The intuition behind the inequality (3.2) is, in essence, the convexity of the involved quantities. One can also establish the upper bound counterpart of (3.2), whence will need to introduce the $\beta^*_\cdot(\cdot;\cdot)$ analogue, where the summation range on the r.h.s. (3.1) is extended to nodal components γ merely *intersecting* $B_x(R)$. However, since the contribution of a single nodal component to the total Betti number is not bounded, and is expected to be *huge* for percolating components, we did not find a useful way to exploit such an upper bound inequality. Instead we are going to seek for a global bound, via Kac-Rice estimating of a relevant local quantity.

Lemma 3.2 (Upper bound). *Let F and i be as in Theorem 1.2. Then*

$$\limsup_{R\to\infty}\frac{\mathbb{E}[\beta_i(R)]}{R^d}<\infty. \tag{3.3}$$

Proof. We use Morse Theory to reduce bounding the expected Betti number $\mathbb{E}[\beta_i(R)]$ from above to a *local* computation, performed with the aid of Kac-Rice method, an approach already exploited by Gayet-Welschinger [11]. Let $\gamma\subseteq\mathbb{R}^d$ be a compact closed hypersurface, and $g:\mathbb{R}^d\to\mathbb{R}$ a smooth function so that its restriction $g|_\gamma$ to γ is a Morse function (i.e. $g|_\gamma$ has no degenerate critical points). Then, as a particular consequence of the Morse inequalities [18, Theorem 5.2 (2) on p. 29], we have

$$b_i(\gamma)\le\mathcal{C}_i(g|_\gamma),$$

where $\mathcal{C}_i(g|_\gamma)$ is the number of critical points of $g|_\gamma$ of Morse index i. Under the notation of Theorem 1.2 it follows that

$$\mathbb{E}[\beta_i(R)]\le\mathbb{E}[\mathcal{C}_i(g|_{F^{-1}(0)\cap B(R)})]\le\mathbb{E}[\mathcal{C}(g|_{F^{-1}(0)\cap B(R)})], \tag{3.4}$$

the r.h.s. of (3.4) being the total number of critical points of g restricted to the nodal set of F lying in $B(R)$, a local quantity that could be evaluated with the Kac-Rice method.

Now we evaluate the r.h.s. of (3.4), where we have the freedom to choose the function g, so long as it is a.s. Morse restricted to $F^{-1}(0)$. As a concrete simple case, we nominate the function

$$\mathbb{R}^d\ni x=(x_1,\dots,x_d)\mapsto g(x)=\|x\|^2=\sum_{j=1}^d x_j^2,$$

or, more generally, the family of functions $g_p=\|x-p\|^2$, $p\in\mathbb{R}^d$, having the burden of proving that for some $p\in\mathbb{R}^d$, the restriction $g_p|_{F^{-1}(0)}$ of g_p to $F^{-1}(0)$ is Morse a.s. For this particular choice of the family g_p, a point $x\in F^{-1}(0)\setminus\{0\}$ is a critical point of g_p, if and only if $\nabla F(x)$ is collinear to $x-p$. Normalising $v_1:=\frac{x-p}{\|x-p\|}$, this is equivalent to $\nabla F(x)\perp v_j$, $j=2,\dots d$, where $\{v_j\}_{2\le j\le d}$ is any orthonormal basis of $v_1^\perp$, and it is possible to make a locally smooth choice for $\{v_j\}_{2\le j\le d}$ as a function of x (or, rather v_1), since $\mathcal{S}^{d-1}$ admits orthogonal frames on a finite partition of the sphere into coordinate patches.

Now, by [18, Lemma 6.3, Lemma 6.5], a critical point $x\in F^{-1}(0)$, of g_p is degenerate, if and only if $p=x+K^{-1}\cdot v_1$, with K one of the (at most $d-1$) principal curvatures of $F^{-1}(0)$ at x in direction v_1, and, by Sard's Theorem [18, Theorem 6.6], given a sample function F_ω, where $\omega\in\Omega$ is a sample point in the underlying sample space Ω, the collection $A_\omega\subseteq\mathbb{R}^d$ of all "bad" p, so that $g_p|_{F^{-1}(0)}$ contains a degenerate critical point is of vanishing Lebesgue measure, i.e.

$$\mu(A_\omega)=0, \tag{3.5}$$

a.s. We are aiming at showing that there exists $p \in \mathbb{R}^d$ so that a.s. $p \notin A_\omega$; in fact, by the above, we will be able to conclude, via Fubini, that μ-almost all p will do (and then, since, by stationarity of F, there is no preference of points in $\mathbb{R}^d$, we will be able to carry out the computations with the simplest possible choice $p = 0$, though the computations are not significantly more involved with arbitrary p). To this end we introduce the set

$$\mathcal{A} = \{(p, \omega) :\ p \in A_\omega\} \subseteq \mathbb{R}^d \times \Omega$$

on the measurable space $\mathbb{R}^d \times \Omega$, equipped with the measure $d\lambda = d\mu(p) d\mathcal{P}r(\omega)$. Since there is no measurability issue here, an inversion of the integral

$$\lambda(\mathcal{A}) = \int\limits_{\mathcal{A}} d\mu(p) d\mathcal{P}r(\omega) = 0,$$

by (3.5), yields that for μ-almost all $p \in \mathbb{R}^d$,

$$\mathcal{P}r\{p \in A_\omega\} = 0. \tag{3.6}$$

The above (3.6) yields a point $p \in \mathbb{R}^d$, so that $g_p|_{F^{-1}(0)}$ is a.s. Morse, and, in particular (3.4) holds a.s. with $g = g_p$; by the stationarity of F, we may assume that $p = 0$, and we take $g = g_0$. Next we plan to employ the Kac-Rice method for evaluating the expected number of critical points of $g|_{F^{-1}(0)}$ as on the r.h.s. of (3.4). Recall from above that, for this particular choice of g, a point $x \in F^{-1}(0) \setminus \{0\}$ is a critical point of g, if and only if $\nabla F(x)$ is collinear to $v_1 = v_1(x) := \frac{x}{\|x\|}$, or, equivalently, $\nabla F(x) \perp v_j$, $j = 2, \dots d$, where $\{v_j\}_{2 \le j \le d}$ is any orthonormal basis of $v_1^{\perp}$.

Let

$$G(x) = (F(x), \langle v_2, \nabla F(x)\rangle, \dots, \langle v_d, \nabla F(x)\rangle) \tag{3.7}$$

be the Gaussian random vector, and $C_G(x)$ its $d \times d$ covariance matrix. That the joint Gaussian distribution of $G(x)$ is non-degenerate, is guaranteed by the axiom ($\rho 3$), since this axiom yields [27, §1.2.1] the non-degeneracy of the distribution of $\nabla F(x)$ (and hence of any linear transformation of $\nabla F(x)$ of full rank), and $F(x)$ is statistically independent of $\nabla F(x)$. By the Kac-Rice formula [2, Theorem 6.3], using the non-degeneracy of the distribution of $G(x)$ as an input, we conclude that for every $\epsilon > 0$

$$\mathbb{E}[\mathcal{C}(g|_{F^{-1}(0) \cap (B(R) \setminus B(\epsilon))})] = \int\limits_{B(R) \setminus B(\epsilon)} K_1(x) dx, \tag{3.8}$$

where for $x \neq 0$, the density is defined as the Gaussian integral

$$K_1(x) = K_{1;F}(x) = \frac{1}{(2\pi)^{d/2} \sqrt{|\det C_G(x)|}} \cdot \mathbb{E}[|\det H_G(x)| \big| G(x) = 0], \tag{3.9}$$

and $H_G(\cdot)$ is the Hessian of G. Next we apply the Monotone Convergence theorem on (3.8) as $\epsilon \to \infty$, upon bearing in mind that $x = 0$ is not a zero of F a.s., we obtain

$$\mathbb{E}[\mathcal{C}(g|_{F^{-1}(0) \cap B(R)})] = \int\limits_{B(R)} K_1(x) dx, \tag{3.10}$$

extending the definition of K_1 at $x = 0$ arbitrarily.

In what follows we are going to show that $K_1(\cdot)$ is *bounded* on $\mathbb{R}^d$, which, in light of (3.10) is sufficient to yield (3.3), via (3.4). To this end we observe that, since F is stationary, the value of K_1 is defined intrinsically as a function of $v_1 \in \mathcal{S}^{d-1}$, no matter how v_j, $j \geq 2$ were determined, as long as they constitute an o.n.b. of $v_1^{\perp}$, i.e.

$$K_1(x) = K_1(x/\|x\|) = K_1(v_1),$$

despite the fact that the law of $G(x)$ does, in general, depend on the choice of the vectors $\{v_j\}$, $j \geq 2$.

The upshot is that, since, given $v_1 \in \mathcal{S}^{d-1}$, one can choose $\{v_j\}_{2\leq j\leq d}$ locally continuously, the law of $G(x)$ is determined in a locally continuous and non-degenerate way as a function of v_1, meaning that $|\det C_G(\cdot)| > 0$. Hence $K_1(\cdot)$ in (3.9) is a continuous function of $v_1 \in \mathcal{S}^{d-1}$, and therefore it is bounded by a constant depending only on the law of F (though not necessarily defined continuously *at* the origin). As it was readily mentioned, the boundedness of K_1 is sufficient to yield the statement (3.3) of Lemma 3.2.

□

The following lemma is a restatement of [25, Proposition 5.2] for random fields satisfying (ρ4), and of [7, Theorem 1.3(i)] for Berry's monochromatic isotropic waves in higher dimensions, and thereupon its proof will be conveniently omitted here.

Lemma 3.3. *Let $F : \mathbb{R}^d \to \mathbb{R}$ be a Gaussian random field, $\mathcal{H}(d-1)$ the collection of all diffeomorphism classes of closed $(d-1)$-manifolds that have an embedding in $\mathbb{R}^d$, and for $H \in \mathcal{H}(d-1)$ denote $\mathcal{N}_{F,H}(R)$ the number of nodal components of F, entirely contained in $B(R)$ and diffeomorphic to H. Then if F either satisfies (ρ4) or it is Berry's monochromatic isotropic waves, one has:*

$$\liminf_{R\to\infty} \frac{\mathbb{E}[\mathcal{N}_{F,H}(R)]}{R^d} > 0.$$

3.2. **Proof of Theorem 1.2.**

Proof. First we aim at proving (1.4), that will allow us to deduce (1.5), with the help of (3.2). Take

$$\eta := \limsup_{R\to\infty} \frac{\mathbb{E}[\beta_i(R)]}{R^d}. \tag{3.11}$$

Then, necessarily $\eta < \infty$ is finite, thanks to Lemma 3.2. We claim that, in fact, (3.11), is a limit, whence it is sufficient to show that

$$\liminf_{R\to\infty} \frac{\mathbb{E}[\beta_i(R)]}{R^d} \geq \eta. \tag{3.12}$$

To this end we take $\epsilon > 0$ to be an arbitrary positive number, and, by the definition of η as a $\limsup$, we may choose $r = r(\epsilon) > 0$ so that

$$\frac{\mathbb{E}[\beta_i(r)]}{r^d} > \eta - \epsilon. \tag{3.13}$$

We now take $R > r$, and appeal to the Integral Geometric sandwich (3.2), so that taking an expectation of both sides of (3.2) yields

$$\mathbb{E}[\beta_i(R)] \geq \frac{1}{\operatorname{Vol} B(r)} \int\limits_{B(R-r)} \mathbb{E}[\beta_i(x;r)]dx = \frac{(R-r)^d}{r^d} \cdot \mathbb{E}[\beta_i(r)], \tag{3.14}$$

by the stationarity of F. Substituting (3.13) into (3.14), it follows that

$$\mathbb{E}[\beta_i(R)] \geq (R-r)^d \cdot (\eta - \epsilon),$$

and hence, dividing by R^d, and taking $\liminf\limits_{R\to\infty}$ (note that r is kept fixed), we obtain

$$\liminf_{R\to\infty} \frac{\mathbb{E}[\beta_i(R)]}{R^d} \geq \eta - \epsilon.$$

Since $\epsilon > 0$ is arbitrary, this certainly implies (3.12), which, as it was mentioned above, implies that η in (3.11) is a limit, a restatement of (1.4) (with $c_i = \frac{\eta}{V_d}$).

Next, having proved (1.4), we are going to deduce the convergence in mean (1.5), this time, assuming the axiom ($\rho 1$), yielding that the action of the translations $\{T_x\}_{x\in\mathbb{R}^d}$ is ergodic, proved independently by Fomin [9], Grenander [12], and Maruyama [17] (see also [27, Theorem 3]). Let $0 < r < R$, and denote the random variable

$$\Psi_i(R,r) = \Psi_i(F;R,r) := \frac{1}{\operatorname{Vol} B(r)} \int\limits_{B(R-r)} \beta_i(x;r)dx, \tag{3.15}$$

so that the Integral Geometric sandwich (3.2) reads

$$\Psi_i(R,r) \leq \beta_i(R), \tag{3.16}$$

and the aforementioned ergodic theorem asserts that, for r fixed, as $R \to \infty$,

$$\frac{1}{\operatorname{Vol} B(R-r)} \Psi_i(R,r) \to \frac{\mathbb{E}[\beta_i(r)]}{\operatorname{Vol} B(r)}$$

in mean (and a.s.), so that we may deduce the same for

$$\frac{1}{\operatorname{Vol} B(R)} \Psi_i(R,r) \to \frac{\mathbb{E}[\beta_i(r)]}{\operatorname{Vol} B(r)}, \tag{3.17}$$

in mean.

Now let $\epsilon > 0$ be arbitrary, and use (1.4), now at our disposal, to choose $r = r(\epsilon)$ sufficiently large (but fixed) so that

$$\left| \frac{\mathbb{E}[\beta_i(r)]}{\operatorname{Vol} B(r)} - c_i \right| < \frac{\epsilon}{3}, \tag{3.18}$$

and also,

$$\left| \frac{\mathbb{E}[\beta_i(R)]}{\operatorname{Vol} B(R)} - \frac{\mathbb{E}[\beta_i(r)]}{\operatorname{Vol} B(r)} \right| < \frac{\epsilon}{4}, \tag{3.19}$$

for the function

$$r \mapsto \frac{\mathbb{E}[\beta_i(r)]}{\operatorname{Vol} B(r)}$$

being Cauchy as $r \to \infty$. Next, use (3.17) in order for the inequality

$$\mathbb{E}\left[\left|\frac{1}{\operatorname{Vol} B(R)}\Psi_i(R,r) - \frac{\mathbb{E}[\beta_i(r)]}{\operatorname{Vol} B(r)}\right|\right] < \frac{\epsilon}{3}, \tag{3.20}$$

to hold, provided that R is sufficiently large (depending on r and ϵ). Note that, thanks to (3.16), we have

$$\begin{aligned}
0 \le \mathbb{E}\left[\left|\frac{\beta_i(R)}{\operatorname{Vol} B(R)} - \frac{1}{\operatorname{Vol} B(R)}\Psi_i(R,r)\right|\right] &= \mathbb{E}\left[\frac{\beta_i(R)}{\operatorname{Vol} B(R)} - \frac{1}{\operatorname{Vol} B(R)}\Psi_i(R,r)\right] \\
= \frac{\mathbb{E}[\beta_i(R)]}{\operatorname{Vol} B(R)} - \frac{1}{\operatorname{Vol} B(R)}\mathbb{E}[\Psi_i(R,r)] &= \frac{\mathbb{E}[\beta_i(R)]}{\operatorname{Vol} B(R)} - \frac{\operatorname{Vol} B(R-r)}{\operatorname{Vol} B(r)\operatorname{Vol} B(R)}\mathbb{E}[\beta_i(r)] \\
= \frac{\mathbb{E}[\beta_i(R)]}{\operatorname{Vol} B(R)} - (1+o_{R\to\infty}(1)) \cdot \frac{\mathbb{E}[\beta_i(r)]}{\operatorname{Vol} B(r)} &< \frac{\epsilon}{3}
\end{aligned} \tag{3.21}$$

for R sufficiently large, by (3.15), the stationarity of F, and (3.19). We consolidate all the above inequalities by using the triangle inequality to write

$$\begin{aligned}
\mathbb{E}\left[\left|\frac{\beta_i(R)}{\operatorname{Vol} B(R)} - c_i\right|\right] &\le \mathbb{E}\left[\frac{\beta_i(R)}{\operatorname{Vol} B(R)} - \frac{1}{\operatorname{Vol} B(R)}\Psi_i(R,r)\right] \\
&+ \mathbb{E}\left[\left|\frac{1}{\operatorname{Vol} B(R)}\Psi_i(R,r) - \frac{\mathbb{E}[\beta_i(r)]}{\operatorname{Vol} B(r)}\right|\right] + \mathbb{E}\left[\left|\frac{\mathbb{E}[\beta_i(r)]}{\operatorname{Vol} B(r)} - c_i\right|\right] < \epsilon,
\end{aligned}$$

by (3.18), (3.20) and (3.21). Since $\epsilon > 0$ was an arbitrary positive number, the mean convergence (1.5) is now established. Finally, we observe that Theorem 1.2c is a direct consequence of Lemma 3.3, as follows. By taking any $H \in \mathcal{H}(d-1)$ with $b_i(H) > 0$ and taking into account the contribution to $\beta_i(R)$ of those nodal components of F, diffeomorphic to H, yielding $\beta_i(R) \ge b_i(H) \cdot \mathcal{N}_{F,H}(R)$, and thus Lemma 3.3 together with (1.4) implies that

$$c_i = \lim_{R\to\infty} \frac{\beta_i(R)}{\operatorname{Vol} B(R)} > 0.$$

Theorem 1.2 is now proved.

□

4. Proof of Theorem 1.5

Let $x \in \mathcal{M}$ be a point as postulated in Theorem 1.5, K_x the corresponding covariance kernel, and F_x the centred Gaussian random field defined by F_x. Recall that $f_{x,L}(\cdot)$, defined in (1.9) on $\mathbb{R}^d$ via the identification $T_x(\mathcal{M}) \cong \mathbb{R}^d$, is the scaled version of f_L, converging in the limit $L \to \infty$, to F_x, with accordance to (1.10). By the manifold structure of $\mathcal{M}$, the exponential map $\exp_x : T_x \to \mathcal{M}$ is a diffeomorphism on a sufficiently small ball $B(r) \subseteq T_x$, with $r > 0$ independent of x. Hence, for every $R > 0$, the diffeomorphism types in $B(R) \subseteq \mathbb{R}^d \cong T_x(\mathcal{M})$ are preserved under the *scaled* exponential map

$$\exp_{x;L} : u \mapsto \exp_x(u/L),$$

provided that L is sufficiently large. In particular, if $\gamma \subseteq B(R)$ is a smooth hypersurface, then for every $0 \le i \le d-1$

$$b_i(\gamma) = b_i(\exp_{x;L}(\gamma)), \tag{4.1}$$

Further, for $r > 0$ sufficiently small $\exp_x$ maps $B(r)$ into the geodesic ball $B_x(r)$, so that, for every $R > 0$, and L sufficiently large, we have

$$\exp_{x;L}(B(R)) = B_x(R/L). \tag{4.2}$$

We can then infer from (4.1) combined with (4.2), that

$$\beta_{f_{x,L};i}(R) = \beta_i(f_L; x, R/L) \tag{4.3}$$

holds for every $R > 0$, $L \gg 0$ sufficiently large. We observe that, by the assumption (1.10) of Theorem 1.5, the Gaussian random fields $\{f_{x,L}\}$ converge in law to the Gaussian random field F_x. That alone does not ensure that one can compare the sample functions $f_{x,L}$ to the sample functions F_x, without *coupling* them in a particular way, (i.e. define both on the same probability space Ω to satisfy some postulated properties). Luckily, such a convenient coupling was readily constructed [27, Lemma 4], and we will reuse it for our purposes. For the sake of the completeness and the reader's convenience we state and prove it in what follows.

Lemma 4.1 ([27, Lemma 4]). *For every $R > 0$ and $\alpha > 0$ there exists a coupling of F_x and $f_{x,L}$ and L_0 sufficiently large so that for all $L > L_0$ we have*

$$\mathbb{E}[\|f_{x,L} - F_x\|_{C^1(B(2R))}] < \alpha. \tag{4.4}$$

Proof. We explicate the details in the sketch of the proof given within [27]. First, we take a sufficiently small parameter $\eta = \eta(R, \alpha) > 0$, determined below, and consider the restriction of both F_x and $f_{x,L}$ to an η-net $\mathcal{X} = \mathcal{X}_{R,\eta} = \{z_k\}_{1\le k\le K} \subseteq B(2R)$, i.e., for every $z \in B(2R)$, $B_z(\eta) \cap \mathcal{X} \ne \emptyset$ and for all $1 \le k \ne l \le K$, $\|z_k - z_l\| \ge \eta$. Our first goal is to continuously couple the multivariate Gaussian vectors

$$V = V(F_x, \mathcal{X}) = (v_k)_{1\le k\le K} := (F_x(z_k))$$

and

$$W_L = W(f_{x,L}, \mathcal{X}) = (w_k)_{1\le k\le K} := (f_{x,L}(z_k)),$$

so that, assuming that both R and η are fixed, for L sufficiently large,

$$\mathbb{E}[\|V - W_L\|_\infty] < \alpha/2, \tag{4.5}$$

where $\|\cdot\|_\infty$ is the usual max norm in $\mathbb{R}^K$. Let $A = A_{K\times K}(F_x; \mathcal{X})$ and $B_L = B_{K\times K}(f_{x,L}; \mathcal{X})$ be the covariance matrices of V and W_L respectively; A is positive semi-definite (though could, in general, be singular), and so is B_L.

Recall that we assumed that (1.10) holds, so that all the entries of B_L converge as $L \to \infty$ to the respective entries of A. If A were nonsingular, then both A and B_L would possess unique positive definite square roots, S_0 and S_L respectively, so that, as $L \to \infty$, the symmetric matrices S_L converge to S_0 in the max norm, by the standard perturbation theory [13]. In that case we could take $Z_{K\times 1}$ to be the K-variate standard Gaussian vector, and set $V := S_0 Z$ with covariance $S_0 S_0^t = S_0^2 = A$ and $W_L := S_L Z$ with covariance $S_L^2 = B_L$, claiming that to be the required coupling of V and W_L satisfying (4.5) for L sufficiently large. The estimate (4.5) holds by the trivial inequality

$$\|V - W_L\|_\infty \le \|V - W_L\|_1,$$

which, in turn, holds thanks to the convergence of S_L to S_0 (recall that R and η are fixed, and so is $\mathcal{X}_{R,\eta}$).

However, since the nonsingularity of the covariance matrix A is not guaranteed (though it does follow from axiom (ρ4), an assumption we are not willing to make here), we will have to add to the said argument. We observe that the only source of the problem is the eigenvectors with vanishing eigenvalue, corresponding to unit eigenvectors with *small* eigenvalues of B_L (so the variance of the corresponding Gaussian random variable is small), whose contribution to (4.5) could be made arbitrarily small, with precise details as follow. Let $\epsilon = \epsilon(F_x, R, \eta) > 0$ be a small parameter (much smaller than η, which is assumed to be fixed). Since both A and B_L are real symmetric, one has an orthogonal diagonalisation $A = P_0^t D_0 P$ and $B_L = P_L^t D_L P_L$ where we may assume with no loss of generality that $D_0 = \operatorname{diag}(\beta_1, \ldots, \beta_r, 0, \ldots, 0)$ with $r = \operatorname{rk}(A)$, for all $k \le r$ we have $\beta_k > 0$, and $D_L = \operatorname{diag}(\gamma_{L;1}, \ldots, \gamma_K)$ with $|\gamma_{L;k} - \beta_k| < \epsilon$ for $1 \le k \le r$, and $0 \le \gamma_{L;k} < \epsilon$ for $r+1 \le k \le K$. We may then define the symmetric matrices

$$S_0 := P_0^t \sqrt{D_0} P_0 \tag{4.6}$$

and

$$S_L := P_L^t \sqrt{D_L} P_L, \tag{4.7}$$

satisfying $S_0^2 = A$, $S_L^2 = B_L$.

Let $\widetilde{D_0} = \operatorname{diag}(\beta_1, \ldots, \beta_r)$, $\widetilde{D_L} = \operatorname{diag}(\gamma_1, \ldots, \gamma_r)$ and $\widetilde{P_0}$ be the $r \times K$ matrix containing the first r rows of P_0 and $\widetilde{P_L}$ be the $r \times K$ matrix containing the first r rows of P_L. Then, by the above, $A = \widetilde{P_0}^t \widetilde{D_0} \widetilde{P}$ and $B = \widetilde{P_L}^t \widetilde{D_L} \widetilde{P_L} + O(\epsilon \cdot K)$ in ∞-norm, so that, since all the entries of B_L converge to those of A, S_0 and S_L in (4.6) and (4.6) respectively satisfy

$$\|S_0 - S_L\|_\infty \to 0$$

as $L \to \infty$, provided we choose $\epsilon > 0$ sufficiently small as a function of L. The coupling $V := S_0 Z$ and $W_L := S_L Z$ of Gaussian random vectors of covariance A and B_L will then satisfy (4.5), provided that we took L sufficiently large.

Our next goal is to couple F_x, $f_{x,L}$, V and W_L so that we have

$$F_x|_{\mathcal{X}} = V \tag{4.8}$$

and $f_{x,L}|_{\mathcal{X}} = W_L$. Let $\mathcal{H}$ be the reproducing kernel Hilbert space, $\mathcal{H}_0 \subseteq \mathcal{H}$ its subspace $\mathcal{H}_0 = \{f \in \mathcal{H} : \forall k \le K. f|_{\mathcal{X}} \equiv 0\}$ of functions vanishing on $\mathcal{X}$ (i.e. on all points z_k), and $\mathcal{H}_1 = \mathcal{H}_0^\perp$ of finite dimension

$$\dim \mathcal{H}_1 = \operatorname{rk}(A) \le K. \tag{4.9}$$

Then one has the orthogonal decomposition $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, the probability measure on $\mathcal{H}$ corresponding to F_x induces probability measures on both $\mathcal{H}_0$ and $\mathcal{H}_1$, and we may write

$$F_x(\cdot) = F_{x;0}(\cdot) + F_{x;1}(\cdot),$$

where $F_{x;0}$ and $F_{x;1}$ are independent random fields based on $\mathcal{H}_0$ and the finite dimensional space $\mathcal{H}_1$. Since $F_{x;0}|_{\mathcal{X}} \equiv 0$, the restriction $F_{x;1}|_{\mathcal{X}}$ is equidistributed with $F_x|_{\mathcal{X}}$, which is, in turn, equidistributed with V. By the dimension considerations (4.9), the equation (4.8) uniquely determines the law of $F_{x;1}$. One can complete a similar procedure for $f_{x,L}$.

All in all, the above steps yield a coupling of F_x and $f_{x,L}$ so that

$$\mathbb{E}[\|F_x|_{\mathcal{X}} - f_{x,L}|_{\mathcal{X}}\|_\infty] < \alpha/2. \tag{4.10}$$

In what follows we prove that the said coupling satisfies the claimed estimate (4.4). For once, the bound

$$\mathbb{E}\left[\|F_x - f_{x,L}\|_{C^0(\overline{B}(2R))}\right] < \alpha \tag{4.11}$$

follows from the standard bound [1, Theorem 1.4.1 on p. 20] on the expected continuity module

$$\omega_F(\eta) := \sup_{\|t-s\|<\eta} |F(t) - F(s)|.$$

The said result with $F = F_x$ and $F = f_{x,L}$, in light of the assumed uniform smoothness (1.11), implies that $\mathbb{E}[\omega_F(\eta)] \to 0$ as $\eta \to 0$, uniformly for L sufficiently large, which, bearing in mind (4.10), is sufficient for (4.11).

Finally, we infer (4.4) from (4.11) and an a priori bound on

$$\|F_x - f_{x,L}\|_{C^2(\overline{B}(2R))},$$

via the usual Landau-Kolmogorov inequality

$$\|G\|_{C^2(\overline{B}(2R))} \le M\|G\|^{1/2}_{C^0(\overline{B}(2R))} \cdot \|\partial_1^2 G\|^{1/2}_{C^0(\overline{B}(2R))},$$

where $G = F_x - f_{x,L}$, and $M = M(d;2,1) > 0$ is an absolute constant. It follows from [1, Theorem 1.4.1], again appealing to the uniform smoothness (1.11) (this time requiring a bound for the 3rd derivatives), that for sufficiently large L, the norm $\|\partial_1^2(F_x - f_{x,L})\|_{C^0(\overline{B}(2R))}$ is uniformly bounded (by a constant depending on R), and, hence (4.4) follows, at the expense of replacing α with a smaller parameter $\alpha' = \frac{\alpha^2}{\sqrt{K}}$ in (4.11).

□

Now we aim to prove the following result, that, taking into account Theorem 1.2 applied on F_x, and (4.3), yields Theorem 1.5 at once. We will denote Ω to be the underlying probability space, where all the random variables are going to be defined, and $\mathcal{P}r$ the associated probability measure.

Proposition 4.2. *Under the assumptions of Theorem 1.5, there exists a coupling of F_x and $\{f_{x,L}\}$ so that for every $R > 0$ and $\delta > 0$ there exists a number $L_0 = L_0(R,\delta) \in \mathcal{L}$ sufficiently big, so that for all $L > L_0$ the following inequality holds outside an event of probability $< \delta$:*

$$\beta_{F_x;i}(R-1) \le \beta_{f_{x,L};i}(R) \le \beta_{F_x;i}(R+1). \tag{4.12}$$

In what follows we are going to exhibit a construction of the small exceptional event from [27], where (4.12) might not hold, prove by way of construction that it is of arbitrarily small probability, and finally culminate, this section with a proof that (4.12) holds outside the exceptional event.

For $R>0$, $L\in\mathcal{L}$, $\alpha>0$ we denote the following "bad" events in Ω:

$$\Delta_1=\Delta_1(R,L,\alpha)=\left\{\|f_{x,L}-F_x\|_{C^1(\overline{B}(2R))}>\alpha\right\},$$

and the "unstable" event

$$\Delta_4=\Delta_4(R,\alpha)=\left\{\min_{y\in\overline{B}(2R)}\max\{|F_x(y)|,|\nabla F_x(y)\}<2\alpha\right\},$$

(with the more technical events Δ_2,Δ_3 unnecessary for the purposes of this manuscript), and then set the exceptional event

$$\Delta=\Delta(R,L,\alpha):=\Delta_1\cup\Delta_4. \tag{4.13}$$

The following bounds for the bad events are due to Nazarov-Sodin [27] (see also [25, 4]).

Lemma 4.3. *There exists a coupling of F_x and $\{f_{x,L}\}$ on Ω, so that the following estimates hold.*

a. [27, Lemma 4], cf. Lemma 4.1 above: *For every $R>0$, $\alpha>0$*

$$\limsup_{L\to\infty}\mathcal{P}r\left(\Delta_1(R,L,\alpha)\right)=0.$$

b. [27, Lemma 5]: *For every $R>0$,*

$$\lim_{\alpha\to 0}\mathcal{P}r\left(\Delta_4(R,\alpha)\right)=0.$$

The following lemma, due to Nazarov-Sodin, shows that if a function has no low lying critical points, then its nodal set is stable under small perturbations.

Lemma 4.4 ([27, Lemmas 6-7], [25, Proposition 6.8])**.** *Let α, $R>1$, and $f:B(R)\to\mathbb{R}$ be a C^1-smooth function on an open ball $B=B(R)\subseteq\mathbb{R}^d$ for some $R>0$, such that for every $y\in B(R)$, either $|f(y)|>\alpha$ or $\|\nabla f(y)\|>\alpha$. Let $g\in C^1(B)$ such that $\sup\limits_{y\in B}|f(y)-g(y)|<\alpha$. Then each nodal component γ of $f^{-1}(0)$ lying in $B(R-1)$ generates a nodal component γ' of g diffeomorphic to γ lying in $B(R)$. Moreover, the map $\gamma\mapsto\gamma'$ between the nodal components of f lying in $B(R-1)$ and the nodal components of g lying in $B(R)$ is injective.*

We are now ready to show a proof of Proposition 4.2.

Proof of Proposition 4.2. Let $R>0$ and $\delta>0$ be given. On an application of Lemma 4.3b we obtain a number $\alpha=\alpha(R,\delta)$ so that

$$\mathcal{P}r(\Delta_4(R,\alpha))<\delta/2,$$

and subsequently, we apply Lemma 4.3a to obtain number $L_0=L_0(R,\delta,\alpha)$ so that for all $L>L_0$,

$$\mathcal{P}r(\Delta_1(R,L,\alpha))<\delta/2.$$

Defining the exceptional event as in (4.13), the above shows that

$$\mathcal{P}r(\Delta) < \delta.$$

We now claim that second inequality of (4.12) is satisfied on $\Omega \setminus \Delta$; by the above this is sufficient yielding the statement of Proposition 4.2, and, as it was previously mentioned, also of Theorem 1.5. Outside of Δ we have both

$$\min_{y \in \overline{B}(2R)} \max\{|F_x(y)|, |\nabla F_x(y)\} > 2\alpha$$

and

$$\|f_{x,L} - F_x\|_{C^1(\overline{B}(2R))} < \alpha$$

for $L > L_0$, and these two also allow us to infer

$$\min_{y \in \overline{B}(2R)} \max\{|f_{x,L}(y)|, |\nabla f_{x,L}(y)\} > \alpha$$

for $L > L_0$. The first inequality of (4.12) now follows upon a straightforward application of Lemma 4.4, with F_x and $f_{x,L}$ taking the roles of f and g respectively, whereas the second inequality of (4.12) follows upon reversing the roles of f and g. Proposition 4.2 is now proved.

□

References

[1] Robert J Adler and Jonathan E Taylor. *Random fields and geometry.* Springer Science & Business Media, 2009.

[2] J.-M. Azaïs and M. Wschebor. *Level sets and extrema of random processes and fields.* John Wiley & Sons, Inc., Hoboken, NJ, 2009. doi: 10.1002/9780470434642. URL https://doi.org/10.1002/9780470434642.

[3] Vincent Beffara and Damien Gayet. Percolation of random nodal lines. *Publications mathématiques de l'IHÉS*, 126(1):131–176, 2017.

[4] Dmitry Beliaev and Igor Wigman. Volume distribution of nodal domains of random band-limited functions. *Probability theory and related fields*, pages 1–40, 2018.

[5] Dmitry Beliaev, Stephen Muirhead, and Igor Wigman. Russo-Seymour-Welsh estimates for the kostlan ensemble of random polynomials. *arXiv preprint arXiv:1709.08961*, 2017.

[6] Dmitry Beliaev, Stephen Muirhead, and Igor Wigman. Mean conservation of nodal volume and connectivity measures for gaussian ensembles. *arXiv preprint arXiv:1901.09000*, 2019.

[7] Yaiza Canzani and Peter Sarnak. Topology and nesting of the zero set components of monochromatic random waves. *Communications on Pure and Applied Mathematics*, 72 (2):343–374, 2019.

[8] Anne Estrade and Julie Fournier. Number of critical points of a gaussian random field: Condition for a finite variance. *Statistics & Probability Letters*, 118:94–99, 2016.

[9] S.V. Fomin. the theory of dynamical systems with continuous spectrum. *Doklady Akad. Nauk*, 67:435–437, 1949.

[10] D. Gayet and J.-Y. Welschinger. Lower estimates for the expected Betti numbers of random real hypersurfaces. *Journal of the London Mathematical Society*, 90(1):105–120, 2014.

[11] Damien Gayet and Jean-Yves Welschinger. Betti numbers of random real hypersurfaces and determinants of random symmetric matrices. *Journal of the European Mathematical Society*, 18(4):733–772, 2016.

[12] Ulf Grenander. Stochastic processes and statistical inference. *Arkiv för matematik*, 1 (3):195–277, 1950.

[13] T. Kato. *Perturbation theory for linear operators*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. ISBN 3-540-58661-X. Reprint of the 1980 edition.

[14] U. Krengel. Ergodic theorems. With a supplement by Antoine Brunel. De Gruyter Studies in Mathematics, 6. Berlin-New York: Walter de Gruyter., 1985.

[15] Manjunath Krishnapur, Pär Kurlberg, and Igor Wigman. Nodal length fluctuations for arithmetic random waves. *Annals of Mathematics*, pages 699–737, 2013.

[16] Pär Kurlberg and Igor Wigman. Variation of the Nazarov–Sodin constant for random plane waves and arithmetic random waves. *Advances in Mathematics*, 330:516–552, 2018.

[17] Gisiro Maruyama. The harmonic analysis of stationary stochastic processes. *Memoirs of the Faculty of Science, Kyushu University. Series A, Mathematics*, 4(1):45–106, 1949.

[18] John Willard Milnor, Michael Spivak, Robert Wells, and Robert Wells. *Morse theory*. Princeton university press, 1963.

[19] Stephen Muirhead. A second moment bound for critical points of planar gaussian fields in shrinking height windows. *arXiv preprint arXiv:1901.11336*, 2019.

[20] F. Nazarov and M. Sodin. Asymptotic laws for the spatial distribution and the number of connected components of zero sets of Gaussian random functions. *Zh. Mat. Fiz. Anal. Geom.*, 12(3):205–278, 2016. doi: 10.15407/mag12.03.205. URL `https://doi.org/10.15407/mag12.03.205`.

[21] X.-X. Nguyen. Ergodic theorems for subadditive spatial processes. *Z. Wahrsch. Verw. Gebiete*, 48(2):159–176, 1979. doi: 10.1007/BF01886870. URL `https://doi.org/10.1007/BF01886870`.

[22] Ferenc Oravecz, Zeév Rudnick, and Igor Wigman. The leray measure of nodal sets for random eigenfunctions on the torus. In *Annales de l'institut Fourier*, volume 58, pages 299–335, 2008.

[23] Changbom Park, Pratyush Pranav, Pravabati Chingangbam, Rien Van De Weygaert, Bernard Jones, Gert Vegter, Inkang Kim, Johan Hidding, and Wojciech A Hellwing. Betti numbers of gaussian fields. *arXiv preprint arXiv:1307.2384*, 2013.

[24] P. Sarnak. Private communication, 2017.

[25] P. Sarnak and I. Wigman. Topologies of nodal sets of random band-limited functions. *Communications on Pure and Applied Mathematics*. doi: 10.1002/cpa.21794. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.21794`.

[26] P. Sarnak and I. Wigman. Topologies of nodal sets of random band limited functions. In *Advances in the theory of automorphic forms and their L-functions*, volume 664 of *Contemp. Math.*, pages 351–365. Amer. Math. Soc., Providence, RI, 2016. doi: 10.1090/

conm/664/13040. URL `https://doi.org/10.1090/conm/664/13040`.

[27] M. Sodin. Lectures on random nodal portraits. In *Probability and statistical physics in St. Petersburg*, volume 91 of *Proc. Sympos. Pure Math.*, pages 395–422. Amer. Math. Soc., Providence, RI, 2016.

[28] Igor Wigman. On the distribution of the nodal sets of random spherical harmonics. *Journal of Mathematical Physics*, 50(1):013521, 2009.

[29] Igor Wigman. Fluctuations of the nodal length of random spherical harmonics. *Communications in Mathematical Physics*, 298(3):787, 2010.

Email address: `igor.wigman@kcl.ac.uk`

Department of Mathematics, King's College London